

APPLICATION FOR
UNITED STATES LETTERS PATENT
SPECIFICATION

Inventor(s): Takashi YANASE and Yoshio NAKAO

Title of the Invention: DEVICE AND METHOD FOR PRESENTING
NEWS INFORMATION

0074827-122700

Field of the Invention

With the recent spread of personal computers and the Internet, even individuals can easily obtain a great amount of information. One method for collecting information using the Internet is an information distribution service using electronic mail. In this service, if a mail address is registered in advance in a distribution source, news and a variety of information are automatically distributed by electronic mail. Therefore, we can stay home and obtain the latest information. Using this information service

5
10
15
20
25

In such a case, as a typical technology for extracting articles interesting for a user, there is a technology for searching for and presenting information meeting search requirements by inputting the search requirements using a keyword, etc., corresponding to information that a user wants to extract.

If in this way, we receive news by electronic mail, a plurality of articles, including a plurality of topics are often distributed at one time. Such mail sometimes also includes information other than news,

such as advertisement, etc. Therefore, if the user simply keeps the mail, efficient information acquisition, such as the acquisition of only articles, including a specific topic is difficult, which is another problem.

Next, when searching for articles, a user must input search requirements to the system, and the user must have a clear intention to search for. In the case of the content filtering, it is necessary for a user to input his/her interesting topics to the system in a form of a keyword, etc., prior to the distribution of articles. Therefore, if a user does not have a clear intention to search for or if a user wants to obtain generally topical articles across a variety of fields, such information cannot be appropriately extracted, which is another problem.

Summary of the Invention

It is one object of the present invention is to provide news information presenting device and method for extracting useful information from news information distributed within a time period specified by a user and presenting the extracted useful information to the user in order to solve the problems described above. It is another object of the present

5 The news information presenting device, which
is one aspect of the present invention, comprises a
news information division unit breaking down news
information in units of articles when extracting useful
information from the news information and presenting
10 the extracted information to a user, a useful article
selection unit selecting useful articles, such as
highly topical articles, etc., from the broken-down
articles and a news information re-editing unit
collecting the selected articles, re-editing the
15 collected articles in a form suited to be presented
to a user and outputting the re-edited information.

20

25

In the above aspect of the present invention,
5 the useful article selection unit can include an article
grouping unit, an article group selection unit and a
typical article selection unit.

The article group selection unit selects one or more groups from the plurality of groups classified by the article grouping unit, based on topicality degree for indicating the topicality of each group.

In this case, the article group selection unit
20 can also calculate the topicality based on both the
total number of articles belonging to a group and the
number of the information sources of articles belonging
to the group.

In the above aspect of the present invention,
25 the news information presenting device can also further

comprise a user instruction input unit for receiving user's instructions on news information presenting. Alternatively, the typical article selection unit can select one or more typical articles according to the user's instructions.

In this case, the typical article selection unit can also select typical articles according to a user's instruction on the distribution date of news information.

In the above aspect of the present invention, the news information re-editing unit can also output one or more typical articles of each of the one or more groups in order based on the topicality and/or distribution date of the typical articles of the one or more selected groups described above.

In this case, the news information re-editing unit can also further output one or more typical articles of each of the one or more groups as related articles in addition to the one or more typical articles of the one or more groups, can also output a simplified typical article obtained by omitting a part of the content of a typical article and can set up a hyper-link to the original typical article in the simplified typical article or can also output graphics included the article of a group belonging to the typical article

In the above aspect of the present invention, the news information re-editing unit can also convert the contents of one or more typical articles of each of the one or more groups into a form suited to be loudly read and output the converted contents in order based on the topicality and/or distribution date of the typical articles of the one or more groups selected by the article group selection unit.

In the news information presenting method, which is one aspect of the present invention, news information is broken down in units of articles, useful articles are selected from the broken-down articles, the selected articles are collected, the articles are re-edited in a form of news information suited to be presented to a user and the re-edited articles are outputted.

In the above aspect of the present invention,
20 useful articles can also be selected by classifying
broken-down articles based on the similarity degree
among the articles into a plurality of groups, selecting
one or more groups from the plurality of groups based
on the topicality degree for indicating the topicality
25 of each group and selecting one or more typical articles

5

10

15

25

form of news information and outputted.

Brief Descriptions of the Drawings

5 The present invention will be more apparent from the following detailed description when the accompanying drawings are referenced.

Fig. 1 shows the basic configuration of the news information presenting device of the present invention.

10 Fig. 2 shows the system configuration of the news information presenting device of the present invention.

Fig. 3 shows the configuration of the news information presenting device.

15 Fig. 4 is a flowchart showing the process of an article division/article information extraction unit (No. 1).

20 Fig. 5 is a flowchart showing the process of an article division/article information extraction unit (No. 2).

Fig. 6 shows an example of the data structure of article information stored in an article information storage unit.

25 Fig. 7 shows a first example of news information distributed by electronic mail.

002227 122700

Fig. 10 shows an example of data stored in the article information storage unit for the news information shown in Figs. 7 and 9 (No. 1).

Fig. 11 shows an example of data stored in the article information storage unit for the news information shown in Figs. 7 and 9 (No. 2).

Fig. 13 is a flowchart showing the process of a related article group generation unit.

Fig. 15 shows the similarity degree between two keywords.

Fig. 17 is a flowchart showing the process of a typical article selection unit.

25 Fig. 19 is a flowchart showing the detailed

5

10

15

20

Fig. 24 shows a fourth example of article re-editing.

Fig. 25 shows an example of news information to be loudly read.

Fig. 26 shows a preferred embodiment in which key related article groups are designated using other information sources.

Fig. 27 shows how to load the program of the present invention into a computer.

Description of the Preferred Embodiments

Fig. 1 shows the basic configuration of the news information presenting device of the present invention. The news information presenting device shown in Fig. 1, extracts, for example, generally topical information from a great amount of news information

distributed, for example, by electronic mail as useful information and presents the extracted information to a user.

5 The news information presenting device 1 of the present invention shown in Fig. 1 comprises a news information division unit 2, a useful article selection unit 3 and a news information re-editing unit 4.

10 The news information division unit 2 breaks down news information distributed, for example, by electronic mail in units of articles.

 The useful article selection unit 3 selects articles useful for a user, such as highly topical articles, etc., from the articles broken down by the news information division unit 2.

15 The news information re-editing unit 4 collects the articles selected by the useful article selection unit 3, re-edits the collected articles in a form suited to be presented to a user and outputs the re-edited articles.

20 According to the configuration described above, since useful articles that are selected from news information distributed, for example, in a form of electronic mail are re-edited in a form of news information and are presented to a user, the user can
25 efficiently obtain useful information.

Fig. 2 shows the system configuration of the news information presenting device of the present invention. Generally, news information is distributed to the news information presenting device 10 shown in Fig. 2 from a plurality of information sources 11a to 11c via a network 12. The distributed news information 13 is provided to the news information presenting device 10. In this case, not only news information distributed via a network 12, for example, like electronic mail, but news information obtained, for example, from a magazine, etc., can also be provided to the news information presenting device 10.

The news information presenting device 10 comprises an article division unit 14 for breaking down news information into individual articles, an article selection unit 15 for selecting the broken-down articles based on user's requirements, etc., an article re-editing unit 16 for re-editing the selected articles in a form of news information, for example, based on user's requirements and a presenting unit 17 for presenting the re-edited news information to the user.

Fig. 3 shows the configuration of the news information presenting device. The news information presenting device 20 shown in Fig. 3 comprises an article division/article information extraction unit

21 corresponding to the article division unit 14 shown in Fig. 2, an article selection unit 15, an article re-editing unit 16, a presenting unit 17 and an article information storage unit 22 for storing the broken-down article information.

The article division/article information extraction unit 21 reads distributed news information 13, extracts the article portions, breaks down the article portions into individual articles, extracts information required to calculate the similarity degree among a plurality of pieces of information/articles, such as the information source, distribution date of each article, etc., and outputs the result to the article information storage unit 22.

The article selection unit 15 includes a related article group generation unit 25, a topicality degree calculation unit 26 and a typical article selection unit 27. The article selection unit 15 reads article information stored in the article information storage unit 22, selects information useful for a user and outputs the result to the article re-editing unit 16.

The related article group generation unit 25 extracts articles distributed, for example, within a time period specified by a user from the article information storage unit 22, calculates the similarity

degree among the distributed articles, groups articles with related contents and generates a related article group.

5 The topicality degree calculation unit 26 calculates the topicality degree for indicating the topicality degree of each of the related article groups generated by the related article generation unit 25. This calculation is described later.

10 The typical article selection unit 27 selects highly topical related article groups based on the topicality degree of each group calculated by the topicality degree calculation unit 26 and further selects typical articles and also related articles, if required, from the selected related article groups.
15 This selections of both typical articles and related articles are described later.

20 The article re-editing unit 16 collects both the typical and related articles selected by the typical article selection unit 27 of the article selection unit 15, determines the output order of the articles, arranges the articles, for example, based on the appropriate form or layout suited for a user's usage and generates news information suited to be presented to the user.

25 The presenting unit 17 presents the news

Fig. 4 is a flowchart showing the process of an article division/article information extraction unit. When in Fig. 4 the process is started, first, in step S1 news information is read, and in step S2 the document format of the news information is analyzed. In this description it is assumed that the document format of news information is already determined, and the document format analysis method is described later.

Fig. 5 is a flowchart showing the process of the article division/article information extraction unit in the case where articles are extracted, for example, from information in a magazine, etc. When Fig. 5 is compared with Fig. 4, only the difference between them

is that in Fig. 5, after in step S1 news information is read, in step S10 text information is extracted and in step S2 the document format of news information is analyzed.

5 When extracting article information from a printed matter, only text information must be extracted from the printed matter by removing information about photographs, drawings, etc. A variety of technologies can be used for this process. For example, information
10 provided in a form of image data can be handled by a character recognition technology using an OCR (Optical Character Reader). Although as described earlier, a document format is assumed to be determined in document format analysis, format conversion to the specific
15 document format, etc., can be included and executed in the extraction process of this text information.

Fig. 6 shows an example of the data structure of article information stored in the article information storage unit 22. Article information about
20 one article is composed of an article identifier (ID), an information source, a distribution date, the title of an article, the text of an article, graphic data, supplementary information and the length of an article.

25 The article ID is a value peculiar to an article, and it is used to identify each article and to relate

the article to the original article. The information source is the name of an information source distributing the article, such as the name of newspaper, etc. The graphic data store data on drawings and charts, if required, when they are included in an article. The supplementary information is data that are not directly related to the text of an article, but that play a supplementary role, such as link information to a related article, for example a URL (Uniform Resource Locator) address, etc.

It is assumed that the document format analysis in step S2 shown in Figs. 4 and 5 is made according to predetermined rules. Examples of the rules are as follows:

- (1) A line where a specific character, such as "....", "_ _", etc., consecutively appears shall be regarded as a separator for separating articles from each other or separating the title of an article from the text of an article.
- (2) A section separated from a subsequent section with only one-spaced line shall be regarded as the title of an article.
- (3) A section separated from a subsequent section shall be regarded as the text of an article.
- (4) A section beginning with "http://" shall be regarded

5

10

15

20

25

as link information, is separated from the text of an article and is stored as supplementary information shown in Fig. 6.

5 Figs. 10 and 11 show results obtained by extracting the titles and text of articles from the two pieces of text shown in Figs. 7 and 9, respectively, and storing the data in the article information storage unit 22. Since, for example, in the case of electronic mail, both a distribution date and information source
10 are described in a header section, they can be easily extracted. Article numbers 05 and 06 attached to two pieces of mail shown in Fig. 9 are unnecessary information and are not stored in the article information storage unit 22.

15 Fig. 12 is a flowchart showing the entire process of the article selection unit 15 shown in Fig. 3. If in Fig. 12 the process is started, first in step S11, for example, a target time period is inputted by a user. The user desires that highly topical articles are
20 presented as useful articles within this target time period.

Then, in step S12, related article groups are generated by the related article group generation unit 25 shown in Fig. 3. In this generation process of related
25 articles, as described later, generally the similarity

degree among all combinations of two articles are calculated for all articles within the target time period designated by a user, and articles with high similarity degree are formed into one related article group.

Then, in step S13, the topicality degree of each related article group is calculated by the topicality degree calculation unit 26 shown in Fig. 3. This calculation of topicality degree is described later.

Lastly, in step S14, useful articles, such as typical articles described later are selected, for example, from ones with high topicality degree out of related article groups by the typical article selection unit 27 and the process is terminated.

Fig. 13 is a flowchart showing the process of the related article group generation unit 25. If in Fig. 13 the process is started, first, in step S16, article information about all articles distributed within a target time period specified by a user are read from the article information storage unit 22, and in step S17, the similarity degree between arbitrary two articles is calculated for all the articles.

The similarity degree S between two articles A_1 and A_2 is calculated according to the following equation.

$$S = \frac{\sum_{i,j} S_{ij}}{\min\{n(A_1), n(A_2)\}} \quad (1)$$

In the above equation,

5

$$S_{ij} = \frac{|W_{ij}|}{\min(|W_i|, |W_j|)} \quad (2)$$

However, in the above equations, w_i and w_j are keywords extracted from articles A_1 and A_2 , respectively, W_{ij} is the longest matching part character string in these keywords, $|w_i|$, $|w_j|$ and $|w_{ij}|$ are the respective length of the two keywords and the longest matching part character string, respectively, and $n(A_1)$ and $n(A_2)$ are the numbers of keywords extracted from the two articles.

The calculation of similarity degree is further described with reference to Figs. 14 and 15. Fig. 14 shows how to extract keywords from two articles. In this example, it is assumed that one keyword is extracted from each of two article titles. Characters that do not directly indicate the content of an article, such as hiragana, parentheses/brackets, punctuation marks, etc., are regarded as boundaries between

(3)

5

15

20

the total number of articles $m=6$ and the number of information sources $n=3$, and if two articles each are distributed from three information sources ($m_1=m_2=m_3=2$), the solution of equation (3) is 18. However, for example, if two, three and one articles are distributed from the first, second and third information sources, respectively, ($m_1=2$, $m_2=3$ and $m_3=1$), the solution is $9 \times \sqrt[3]{6}$. Therefore, topicality degree in the case where articles are equally distributed is higher than that in the case where articles are not equally distributed.

Fig. 17 is a flowchart showing the process of the typical article selection unit 27 shown in Fig. 3. If in Fig. 17 the process is started, first, in step S24, out of related article groups, one with the highest topicality is selected as a key related article group. In this selection, for example, the threshold value of topicality degree is predetermined, and if a related article group with topicality degree exceeding the threshold value is selected as a key related article group. This threshold value is also relative. Alternatively, related article groups can be arranged in descending order of topicality degree, and a specific number of higher-order related article groups can be selected as key related article groups.

Then, in step S25, a typical article is selected for each selected key related article group and the process is terminated. Out of the article group belonging to the key related article group, a typical article is selected, for example, based on user's requirements. For example, if a user mainly requests the latest information, an article with the latest distribution date is selected as a typical article. If a user requests progress information about how the situation of a specific topic has been going on, an article with a fairly old distribution date, such as an article one week before, is selected as a typical article. If a user requests detailed information about a specific topic, an article on the specific topic with the longest article text is selected as a typical article. If a user requests brief information about a specific topic, an article on the specific topic with the shortest text is selected as a typical article. The number of typical articles for each key related article group is not limited to one, and a plurality of articles can be selected as typical articles.

Fig. 18 is a flowchart showing the entire process of the article re-editing unit 16 shown in Fig. 3. If in Fig. 18, the process is started, first, in step S27, the output order of the article groups selected by the

typical article selection unit 27 of the article selection unit 15 is determined. For example, these typical articles can be arranged in descending order of the topicality degree for each key related article group or in descending order of the distribution date.

Then, in step S28, the article information of the selected article group is read from the article information storage unit 22, in step S29, the read article information is processed, if required, for example, based on user's requirements, the results are arranged according to both the output order and layout suited for a user's usage and are outputted to the presenting unit 17. Then, the process is terminated.

Fig. 19 is a flowchart showing the detailed article information locating process in step S29 shown in Fig. 18. If in Fig. 19 the process is started, first, in step S31 titles and a plurality of pieces of text to be arranged are extracted from the article information read in step S28, in step S32 the titles are located, and in step S33 it is judged whether the plurality of pieces of text should be displayed. If it is judged that the plurality of pieces of text are not displayed, in step S36 it is judged whether there is another article to be located. If it is judged that there is another article to be displayed, processes

in steps S31 and after are repeated. If it is judged that there is no other article, the process is terminated without any process.

5 If in step S33 it is judged that the plurality of pieces of text are also displayed, in step S34 the plurality of pieces of text are processed based on user's requirements, in step S35 the processed plurality of pieces of text are located and the processes in steps 36 and after are executed. Necessary
10 respective processes are executed for the plurality of pieces of text depending on the user's requirements, that is, whether full text, a part in the beginning or a summary should be displayed.

Fig. 20 shows an example of requirement input
15 screen for inputting user's requirements on news information presenting. Although in Fig. 2 it is assumed that user's requirements are provided to the article selection unit 15, in Fig. 3 these requirements can be provided to the related article group generation
20 unit 25 in such a way to input a target time period in step S11 shown in Fig. 12 or can also be directly provided to the article re-editing unit 16 without passing via the article selection unit 15.

In the example shown in Fig. 20, it is requested
25 that the target time period is October 1 to October

31, that the number of topics to be displayed, that is, the number of related article groups is three, that two articles should be selected in descending order of distribution date for each topic as typical articles and that articles should be displayed in a form of summarized article text.

Figs. 21 through 24 show examples of re-edited articles. In the first example shown in Fig. 21, only one typical article is selected from each key related article group, and the title and text of the article are listed. The information sources and distribution dates of the articles are also outputted.

In the second example shown in Fig. 22, a plurality of typical articles are selected from one key related article group. For example, an optimal article meeting user's requirements is outputted as a typical article, and other typical articles are outputted as related articles. If a typical article includes a drawing, chart, etc., which is not shown in Fig. 22, the graphic data can also be read from the article information storage unit 22 and can also be outputted together with article text.

If the content of an article is outputted, the full text of the article can be outputted. If a user wants to output in a brief form or information amount

to be outputted on a display screen is restricted, only a part of the text can be also outputted. In this case, for example, only a predetermined amount of information in the beginning of the text of an article can be outputted or the summary of the article can be generated and outputted using both the title and text of the article. For the generation of such a summary, see the following reference.

Japanese Patent Laid-open No. 10-207891, Yoshio Nakao, "Document Summarizing Device and Its Method"

In the third example shown in Fig. 23, only a part of the text of an article is extracted or a summary is generated and outputted. In this case, as underlined and shown in Fig. 23, the full text of an original article can be accessed by setting a hyperlink to each title of articles. This can be implemented by a publicly known technology, such as an HTML (Hypertext Markup Language), etc.

In the fourth example shown in Fig. 24, one typical article and the titles of related articles as other typical articles are displayed on the left and right sides for each key related article group, only a part of article text is also outputted and further hyperlinks are set to the titles of articles.

In this way, a method for outputting related

article information, a method for outputting only a part of article text, a method for outputting graphics and a method for setting a hyperlink can also be arbitrarily combined.

5 Furthermore, news information can also be re-edited in a form suited for loud reading as one process of the article re-editing unit 16. Fig. 25 shows an example of the result of re-editing an article in a form suited for such loud reading. In the case of
10 loud reading, it is preferable for a content to be outputted to be as simple as possible. Therefore, in this case, as shown in Fig. 25, for example, only the titles of articles are laid out.

15 Lastly, the presenting unit 17 shown in Fig. 3 actually presents news information re-edited by the article re-editing unit 16 to a user. In this case, the re-edited news information can be displayed on the screen, the information can be printed using a printer or further the audio output of output text for loud
20 reading can be used together with the display on the screen by a speech synthesis technology.

Fig. 26 shows a different preferred embodiment. In this different preferred embodiment, if a key related article group is selected from article groups generated
25 by the related article group generation unit 25 shown

in Fig. 3, the key related article group is determined by another information source without calculating topicality degree. For example, the respective similarity degrees between each of the titles of the typical articles selected in advance from each related article group based on user's requirements and the title of an article in another journal is calculated, the similarity degrees are compared and a related article group belonging to a typical article with high similarity degree is designated as a key related article group.

In this example, not only the comparison between the title of each typical article and the first title out of the article titles of another journal is made, but the comparison between the title of each typical article and the last title of another journal can also be made. In this way, a related article group to which a typical article at the bottom, "Electronics Show 1999 opened" belongs is also designated as a key related article group.

As the last preferred embodiment, the loading of a program used in the present invention into a computer is described. For example, although the news information presenting device shown in Fig. 3 is generally implemented, for example, by a personal

computer installed near a user, it can also be implemented by a computer installed on the side of a news information presenting service operator. For example, news information can be presented by enabling
5 a computer to execute programs specified in the claims 19 and 20 of the present invention, programs described in flowcharts shown in Figs. 4, 5, 12, 13 and 16 through 19, etc.

Fig. 27 shows how to load such a program into
10 a computer. The computer 51 for implementing the news information presenting device shown in Fig. 3 comprises a main body 52 and a memory 53. For the memory 53, a random-access memory (RAM), a hard disk, a magnetic disk, etc., can be used. A program is stored in such
15 a memory, and news information can be presented to a user by the main body 52 executing the program.

The news information presenting of the present invention can also be conducted by loading such a program into the computer 51 from a program provider
20 via a network 54. In this case, transmission signals obtained by modulating a carrier wave with data signals representing the program is transmitted from the program provider to the computer 51 via the network 54, which is a transmission medium, and on receipt of
25 the transmission signals, the computer 51 demodulates

and reproduces the program. In this way, the program can be executed.

Furthermore, such a program can also be stored in a portable storage medium on the market, and the news information presenting of the present invention can also be conducted by loading the portable storage medium 55 into the main body 52. The portable storage medium 55 can be sold and bought on the market. For example, for the memory, a variety of storage media, such as a floppy disk, a CD-ROM, an optical disk, a magneto-optical disk, etc., can be used.

As described in detail above, according to the present invention, distributed news information is broken down into individual articles, articles with related contents are grouped as a related article group, and the news information is re-edited and presented. Therefore, a user can efficiently obtain useful information. Topicality degree is calculated for article groups with related contents, and only article groups with high topicality degree can be presented to a user. Therefore, if a user has no clear intention to search or if a user wants to know generally topical information regardless of a field, useful information can be obtained. This greatly contributes to the improvement of the utility of news information

presenting devices.

[illegible]